Causality in Dynamical Systems

Alan Barzilay†
Advisors: Paulo Inácio Prado∗ , Renato Mendes Coutinho⋆
Curso de Ciências Moleculares - USP†

Instituto de Biociências - USP∗

Centro de Matemática, Computação e Cognição - UFABC⋆

Introduction

When analyzing causality between two events, it may be tempting to assume they are causally related when they are correlated, but correlation does not imply causation. Correlation studies are useful to create new hypothesis and explore the data, but may be misleading for a small data set and for complex or non-linear dynamics.

To avoid mirage correlations and problems as such, we can study the causality directly with the Convergent Cross Mapping technique. This approach comes from the idea that time series from two distinct variables are causally related if they are coupled and part of the same dynamical system.

Takens Theorem [1]

"Consider a diffeomorphism $\phi : \mathbf{M} \to \mathbf{M}$ on some compact manifold $\mathbf{M}$ of dimension m, along with $2m + 1$ observation functions $y_k : \mathbf{M} \to \mathbb{R}$, smoothly; by "smooth" we mean at least $\mathbb{C}^2$. Restrict the y_k to have the lag relationships corresponding to a collection of sets $\mathbf{C}$ and lags $\mathbf{b}$ under the dynamical system ϕ, and impose the following generic properties on ϕ :

1. The set $\mathbf{A}$ of periodic points with period $\mathbf{p} < \mathbf{max}(\mathbf{b_k})$ has finitely many points,

2. The eigenvalues of $(\mathbf{D}\phi^{\mathbf{b}})_{\mathbf{x}}$ at each $\mathbf{x}$ in a compact neighborhood $\mathbf{A}$ are distinct and not equal to 1.

Then, for generic $< y_k > \in \tilde{Y}_{C,b}^{2m+1}$, the mapping described by:

$$\mathbf{\Phi}_{(\phi,<y_r>)} = (y_1(\mathbf{x}), y_2(\mathbf{x}), ..., y_{2m+1}(\mathbf{x}))$$
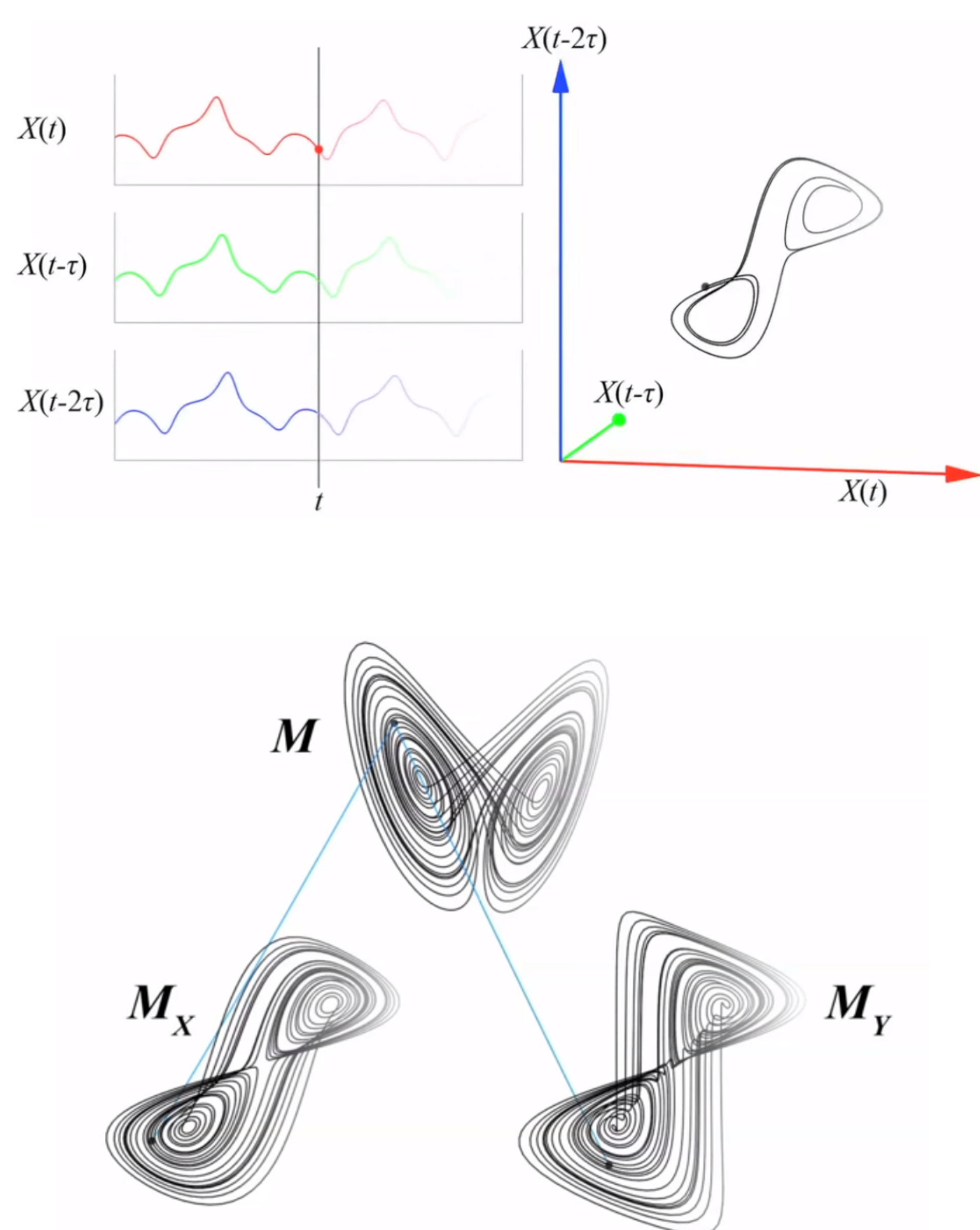
is an embedding."

Figure 1: Reconstruction of the shadow manifold from the time series of lagged coordinates of the Lorenz attractor [2].

Convergent Cross Mapping

The Takens theorem guarantees us that by analyzing the time series of one of the projections of the manifold it is possible to reconstruct a "shadow manifold" that maintains its topological properties. So if we have the state variables Y and X, whom are coupled and participate in the same dynamical system, then both reconstructions are a mapping one to one of the original manifold and subsequently from each other.

This is the basis of the technique known as "Convergent Cross Mapping", where we utilize the Takens theorem to reconstruct the shadow manifold of the state variables to analyze causal relations of the systems trough the time series. By finding the nearest neighbors of a point in M_x at the time t, it can estimate another point in M_y trough a simplex projection and study the correlation between the estimate and the real values of Y.
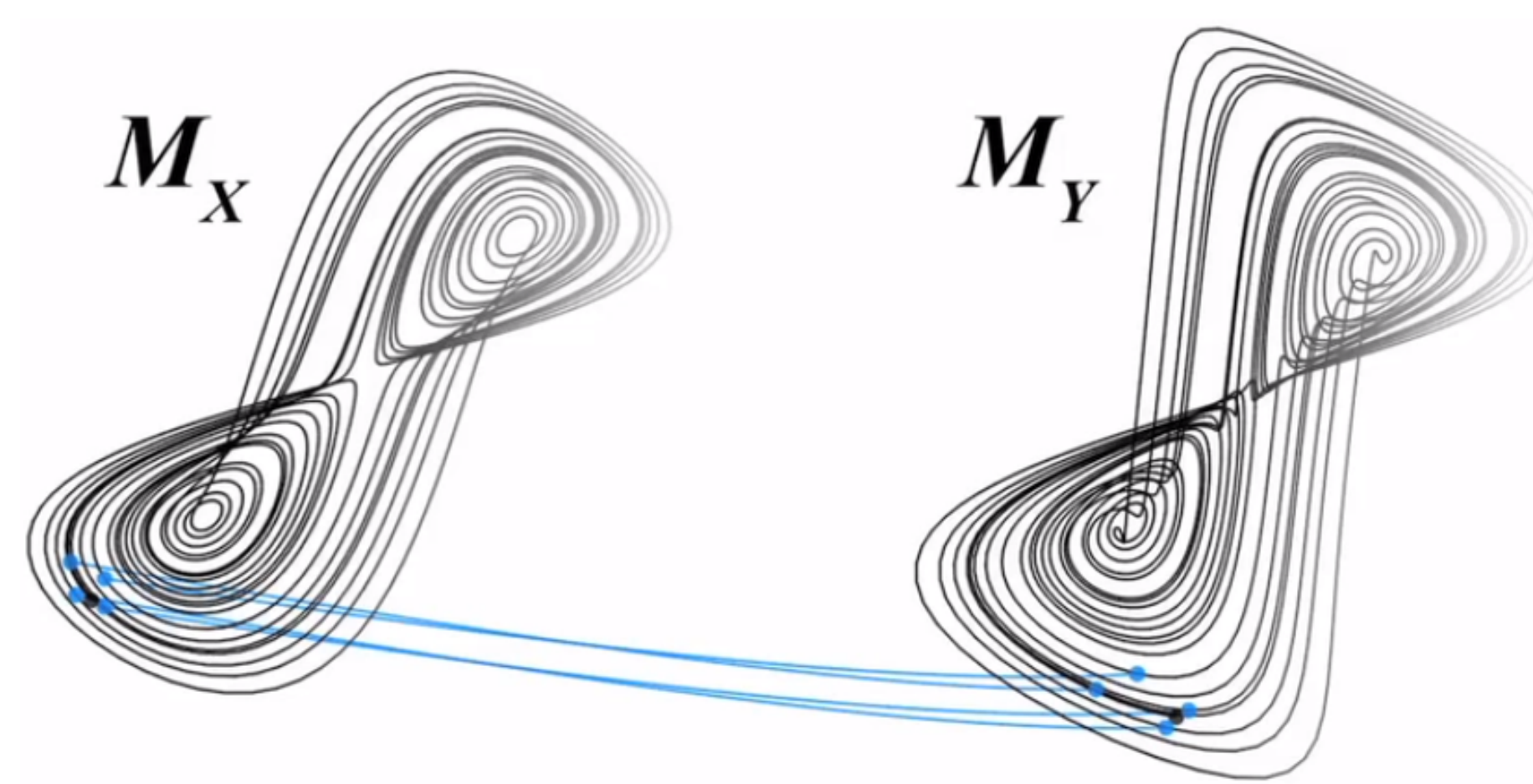

Figure 2: Mapping one to one of the nearest neighbors in M_x to M_y [2].

Simplex Projection

Given the nearest neighbors of a certain point Y_t, we can create a simplex. The simplex projection works as an weighted mean of the distances between the point Y_t and the vertices of these simplex. To predict the new point Y_{t+1} we utilize the relative distance between the vertices at this new moment to project the location of Y_{t+1}
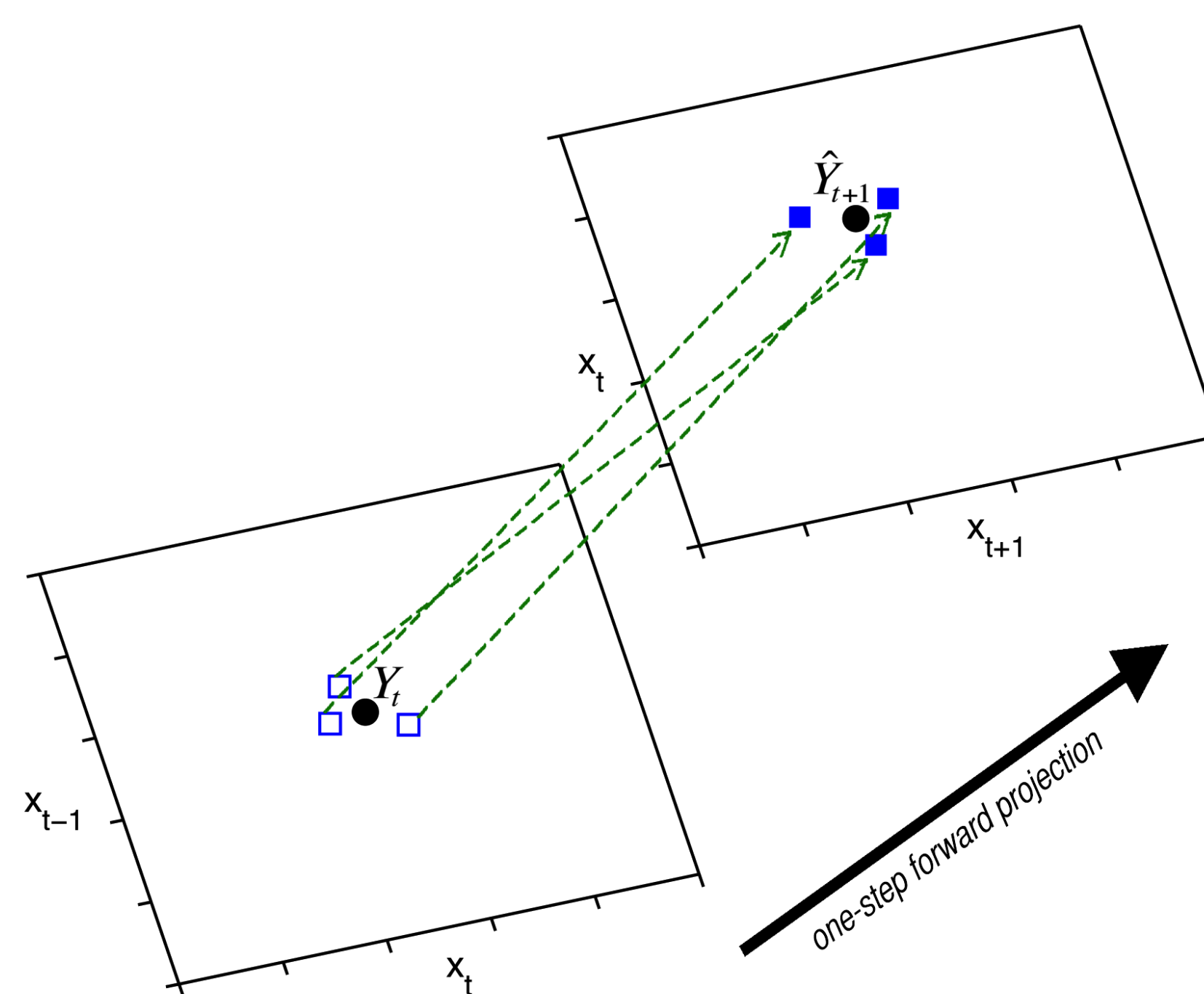

Figure 3: Simplex projection in 2D [3].

S-map

S-maps are an extension of standard linear autoregressive models in which the coefficients depend on the location of the predictee Y_t in an E-dimensional embedding.

From a library set X, we calculate the new coefficients for each new prediction through singular value decomposition. Each vector in the library receives a weight proportional to its distance to the predictee . The weight is determined by a parameter called θ in the following manner:

$$\omega(d) = e^{-\frac{\theta d_{it}}{\bar{d}}}$$

Where $\bar{d}$ is the average distance between neighbors in the library and d_{it} is the distance between Y_t and the ith neighbor vector X_i in the library embedding. Nonlinearity increases as θ increases and in the special case when $\theta = 0$ we have a linear model.
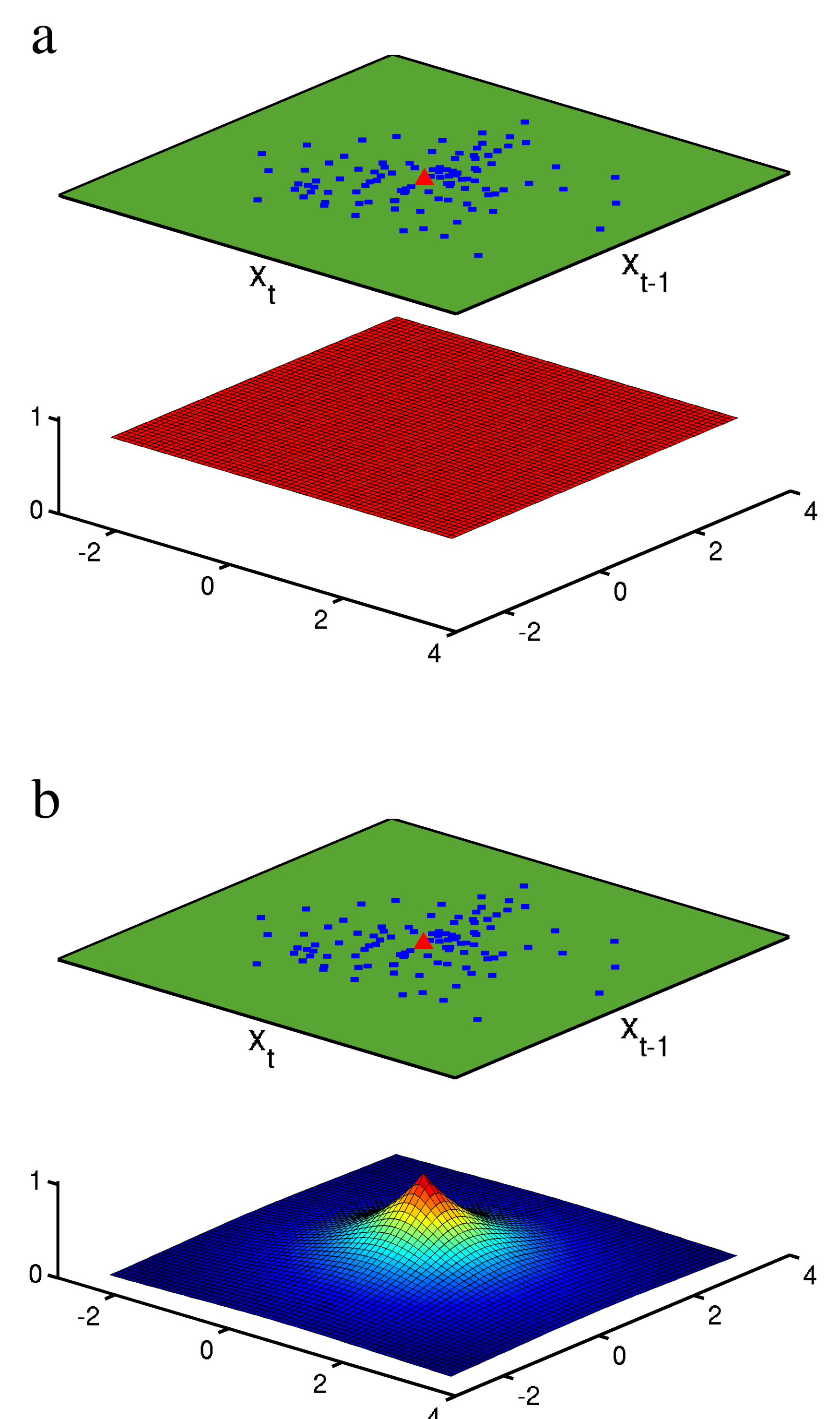

Figure 4: Examples illustrating the S-map procedure for a linear map (a) and a nonlinear map (b) with embedding dimension equal to two (figure and caption from [3]).

Next Steps

In my next steps I intend to use these techniques to analyze time series and find causal relations between the variables of the system. The applications of these methods in biology are truly diverse, ranging from epidemiology to paleontology.

Bibliography

[1] Deyle ER, Sugihara G (2011) *Generalized Theorems for Nonlinear State Space Reconstruction.*

[2] `http://deepeco.ucsd.edu/video-animations/`

[3] Sugihara G (2014) *Distinguishing random environmental fluctuations from ecological catastrophes for the North Pacific Ocean.* Nature 435, 336–340 doi:10.1038/nature03553